

1600

1600

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/08/939,905D

DATE: 11/16/2001  
TIME: 10:20:14

Input Set : A:\Sequence Listing.txt  
Output Set: N:\CRF3\11162001\H939905D.raw

**ENTERED**

3 <110> APPLICANT: Gijzen, Mark  
5 <120> TITLE OF INVENTION: Soybean Seed Coat Peroxidase Structural Gene And Regulatory

Region

7 <130> FILE REFERENCE: 76-105  
9 <140> CURRENT APPLICATION NUMBER: US 08/939,905D  
C--> 11 <141> CURRENT FILING DATE: 1997-09-29  
13 <150> PRIOR APPLICATION NUMBER: US 08/723,414  
15 <151> PRIOR FILING DATE: 1996-09-30  
17 <160> NUMBER OF SEQ ID NOS: 20  
19 <170> SOFTWARE: PatentIn version 3.0

21 <210> SEQ ID NO: 1  
23 <211> LENGTH: 1244

25 <212> TYPE: DNA

27 <213> ORGANISM: Glycine max

29 <220> FEATURE:

31 <221> NAME/KEY: CDS

33 <222> LOCATION: (1)..(1056)

35 <220> FEATURE:

37 <221> NAME/KEY: sig\_peptide

39 <222> LOCATION: (1)..(78)

42 <400> SEQUENCE: 1

43	atg ggt tcc atg cgt cta tta gta gtg gca ttg ttg tgt gca ttt gct	48
44	Met Gly Ser Met Arg Leu Leu Val Val Ala Leu Leu Cys Ala Phe Ala	
45	1 5 10 15	
46	atg cat gca ggt ttt tca gtc tct tat gct cag ctt act cct acg ttc	96
47	Met His Ala Gly Phe Ser Val Ser Tyr Ala Gln Leu Thr Pro Thr Phe	
48	20 25 30	
49	tac aga gaa aca tgt cca aat ctg ttc cct att gtg ttt gga gta atc	144
50	55 60	
51	53 35 40 45	
52	ttc gat gct tct ttc acc gat ccc cga atc ggg gcc agt ctc atg agg	192
53	Phe Asp Ala Ser Phe Thr Asp Pro Arg Ile Gly Ala Ser Leu Met Arg	
54	50 55 60	
55	55 60	
56	55 60	
57	50 55 60	240
58	55 60	
59	55 60	
60	55 60	
61	65 70 75 80	
62	65 70 75 80	
63	65 70 75 80	288
64	65 70 75 80	
65	65 70 75 80	
66	65 70 75 80	
67	65 70 75 80	336
68	65 70 75 80	
69	65 70 75 80	
70	65 70 75 80	
71	65 70 75 80	384
72	65 70 75 80	
73	65 70 75 80	
74	65 70 75 80	432
75	65 70 75 80	
76	65 70 75 80	
77	65 70 75 80	

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/08/939,905D

DATE: 11/16/2001  
TIME: 10:20:14

Input Set : A:\Sequence Listing.txt  
Output Set: N:\CRF3\11162001\H939905D.raw

79 gtt cca tta gga aga agg gac agc tta aca gca aac cga acc ctt gca	480
80 Val Pro Leu Gly Arg Arg Asp Ser Leu Thr Ala Asn Arg Thr Leu Ala	
81 145 150 155 160	
83 aat caa aac ctt cca gca cct ttc ttc aac ctc act caa ctt aaa gct	528
84 Asn Gln Asn Leu Pro Ala Pro Phe Phe Asn Leu Thr Gln Leu Lys Ala	
85 165 170 175	
87 tcc ttt gct gtt caa ggt ctc aac acc ctt gat tta gtt aca ctc tca	576
88 Ser Phe Ala Val Gln Gly Leu Asn Thr Leu Asp Leu Val Thr Leu Ser	
89 180 185 190	
91 ggt ggt cat acg ttt gga aga gct cgg tgc agt aca ttc ata aac cga	624
92 Gly Gly His Thr Phe Gly Arg Ala Arg Cys Ser Thr Phe Ile Asn Arg	
93 195 200 205	
95 tta tac aac ttc agc aac act gga aac cct gat cca act ctg aac aca	672
96 Leu Tyr Asn Phe Ser Asn Thr Gly Asn Pro Asp Pro Thr Leu Asn Thr	
97 210 215 220	
99 aca tac tta gaa gta ttg cgt gca aga tgc ccc cag aat gca act ggg	720
100 Thr Tyr Leu Glu Val Leu Arg Ala Arg Cys Pro Gln Asn Ala Thr Gly	
101 225 230 235 240	
103 gat aac ctc acc aat ttg gac ctg agc aca cct gat caa ttt gac aac	768
104 Asp Asn Leu Thr Asn Leu Asp Leu Ser Thr Pro Asp Gln Phe Asp Asn	
105 245 250 255	
107 aga tac tac tcc aat ctt ctg cag ctc aat ggc tta ctt cag agt gac	816
108 Arg Tyr Tyr Ser Asn Leu Leu Gln Leu Asn Gly Leu Leu Gln Ser Asp	
109 260 265 270	
111 caa gaa ctt ttc tcc act cct ggt gct gat acc att ccc att gtc aat	864
112 Gln Glu Leu Phe Ser Thr Pro Gly Ala Asp Thr Ile Pro Ile Val Asn	
113 275 280 285	
115 agc ttc agc agt aac cag aat act ttc ttt tcc aac ttt aga gtt tca	912
116 Ser Phe Ser Ser Asn Gln Asn Thr Phe Phe Ser Asn Phe Arg Val Ser	
117 290 295 300	
119 atg ata aaa atg ggt aat att gga gtg ctg act ggg gat gaa gga gaa	960
120 Met Ile Lys Met Gly Asn Ile Gly Val Leu Thr Gly Asp Glu Gly Glu	
121 305 310 315 320	
123 att cgc ttg caa tgt aat ttt gtg aat gga gac tcg ttt gga tta gct	1008
124 Ile Arg Leu Gln Cys Asn Phe Val Asn Gly Asp Ser Phe Gly Leu Ala	
125 325 330 335	
127 agt gtg gcg tcc aaa gat gct aaa caa aag ctt gtt gct caa tct aaa	1056
128 Ser Val Ala Ser Lys Asp Ala Lys Gln Lys Leu Val Ala Gln Ser Lys	
129 340 345 350	
131 taaaccaata attaatgggg atgtgcata tagctagcat gtaaaggcaa attaggttgt	1116
133 aaacctttt gctagctata ttgaaataaa ccaaaggagt agtgcata tcaattcgat	1176
135 tttgccatgt acctcttggaa atattatgtataattatttt gaatctcttt aaggtactta	1236
137 attaatca	1244
141 <210> SEQ ID NO: 2	
143 <211> LENGTH: 4700	
145 <212> TYPE: DNA	
147 <213> ORGANISM: Glycine max	
151 <220> FEATURE:	
153 <221> NAME/KEY: promoter	

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/08/939,905D

DATE: 11/16/2001  
TIME: 10:20:14

Input Set : A:\Sequence Listing.txt  
Output Set: N:\CRF3\11162001\H939905D.raw

155 <222> LOCATION: (1)..(1532)  
159 ~220> FEATURE:  
161 <221> NAME/KEY: sig\_peptide  
163 ~222> LOCATION: (1533)..(1610)  
167 ~220> FEATURE:  
169 <221> NAME/KEY: exon  
171 <222> LOCATION: (1533)..(1751)  
175 <220> FEATURE:  
177 <221> NAME/KEY: exon  
179 <222> LOCATION: (2383)..(2574)  
183 <220> FEATURE:  
185 <221> NAME/KEY: exon  
187 <222> LOCATION: (3605)..(3769)  
191 ~220> FEATURE:  
193 <221> NAME/KEY: exon  
195 <222> LOCATION: (4033)..(4515)  
199 <220> FEATURE:  
201 <221> NAME/KEY: Intron  
203 <222> LOCATION: (1752)..(2382)  
207 <220> FEATURE:  
209 <221> NAME/KEY: Intron  
211 <222> LOCATION: (2575)..(3604)  
215 <220> FEATURE:  
217 <221> NAME/KEY: Intron  
219 <222> LOCATION: (3770)..(4032)  
222 <220> FEATURE:  
224 <221> NAME/KEY: CDS  
226 <222> LOCATION: (1533)..(1751)  
230 <220> FEATURE:  
232 <221> NAME/KEY: CDS  
234 <222> LOCATION: (2383)..(2574)  
238 <220> FEATURE:  
240 <221> NAME/KEY: CDS  
242 <222> LOCATION: (3605)..(3769)  
246 <220> FEATURE:  
248 <221> NAME/KEY: CDS  
250 <222> LOCATION: (4033)..(4512)  
254 <400> SEQUENCE: 2 60  
255 tagataaaaa aatggatataatccatc agatgttgg tataactgttt ttttaatcag 60  
257 aattaaaaatt cctcttaat tatcgacata atttttttt gtatatattt tcgcacataat 120  
259 tatttaatac aaatttttat tgcacataga atgtatattt caattttat attggagaac 180  
261 agtacgaaaa cataaaaaaa ctgttatttag aagaaaaaaa tataatgaaa agtttagcta 240  
263 catatatttag ctaaatttagt ttttcttaatt ggctatataa accctattgt actctttgtt 300  
265 atctcacctt tttcatttaa atacatttc accttttaag ttcttatattt tctctcaatt 360  
267 ttcttcgata aaccatgaaa tttaacatgg tataatcagcg ataccaccca ctttggaaagc 420  
269 catgtatggc tagtatggc agccaaaatt tgccctgggtt caagcaaagc aagtgtttat 480  
271 atagatgtga cttttgtga ggaactcatg ccaatggtac tgatttgaa actgagaaaa 540  
273 ctaatttgga gaatttgaat tatgtcatt aaataactcct ctcctgacta cttcgtccc 600  
275 tcaaatttgc accatcatta ttcccaaaaa atttgattac aatgcactaa ttaatgaatg 660

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/08/939,905D

DATE: 11/16/2001  
TIME: 10:20.14

Input Set : A:\Sequence Listing.txt  
Output Set: N:\CRF3\11162001\H939905D.raw

277 tttcttacat tatcatatta tcatatctga cattttgtt ttactttta taataattat	720
279 ttaaaaaagt catacatgca aataattttt taatagttt cagttaaatt tttacagtaa	780
281 aaatgcataa aaattaaact ttattttcc aagtcatcat tttagtcaat cccaaaacaa	840
283 tgattatttt ttgcaaatga atgtttattt aacattaaa tgtagcctaa ttaattctgg	900
285 ttatgggtc aatgtccaa aacctaalgc aagatcttag caagtagata catagatcta	960
287 atttaaact tatcttacg caagagataa aagattata catctagtt taaacattaa	1020
289 ctttggttt tgtgttaaaa aacagtaaca ttttcttaat ttgttagagt gacgtgctcc	1080
291 aaccatatta acgaagattt taattggat tcaagttcat gaacttagta aataagttt	1140
293 ggtcttcagt ttcaattttt cattacaaca tttatgtaaa atatcaacgt tttctgaaat	1200
295 ttgttgctt tggtctccaa ccacattaa gagattatag aaattaattt tcaagaagat	1260
297 aatgattctt actcttgcgt gccctaccat agtacaataa atccactcat aaatcaacaa	1320
299 gtcgtcgta taggaattt ggcattcatat cataaacaat acgtacgtg tattatctag	1380
301 tgtctctca taggaattt gagaattat ttttcttaa aaaaagttaa ttaataaaaa	1440
303 catttgcgtt accgtgagtt acaagaaatc cggccaaattc atctctataa ataaaaggat	1500
305 cttatgaga ggtaaaatca tattactca aa atg ggt tcc atg cgt cta tta	1553
306 Met Gly Ser Met Arg Leu Leu	
307 1 5	
309 gta gtg gca ttg ttg tgc ttt gct atg cat gca ggt ttt tca gtc	1601
310 Val Val Ala Leu Leu Cys Ala Phe Ala Met His Ala Gly Phe Ser Val	
311 10 15 20	
313 tct tat gct cag ctt act cct acg ttc tac aga gaa aca tgt cca aat	1649
314 Ser Tyr Ala Gln Leu Thr Pro Thr Phe Tyr Arg Glu Thr Cys Pro Asn	
315 25 30 35	
317 ctg ttc cct att gtg ttt gga gta atc ttc gat gct tct ttc acc gat	1697
318 Leu Phe Pro Ile Val Phe Gly Val Ile Phe Asp Ala Ser Phe Thr Asp	
319 40 45 50 55	
321 ccc cga atc ygg gcc agt ctc atg agg ctt cat ttt cat gat tgc ttt	1745
322 Pro Arg Ile Gly Ala Ser Leu Met Arg Leu His Phe His Asp Cys Phe	
323 60 65 70	
325 gtt caa gtacgtactt tttttttcc ttccaaatg ccctgcataat ttaacaagat	1801
326 Val Gln	
329 tgctttttc acctagaaaa atgtgtttt ttcaacgatc ttacgtacgt ttgtttgggtt	1861
331 tggaaaaataa atcagaaaa gatcaagaaa atagcttagaa agaaagcaac gttttttaa	1921
333 aaggattttt gtgtgagaaa aatattaaa ctgaagagaa agaaattaaa taagctttc	1981
335 ttgaatgata ttacatgtc ttattactt aaagtcaccc tttttctta agttgtgctt	2041
337 gaagaaaaaa gatgttttc agtttagtt tgattatgc taattatatt ttaattat	2101
339 taattatatac tatatatcta ttaccatata taattattac tatatttcat gatgacaaca	2161
341 gacaagtatt ctaaagaggt atcggtagat gattaatttt ttataaaaa aatctttgc	2221
343 gtgtatagat attctttat aattggtgca gaaacttgta atgctattt caattatct	2281
345 tacattgatt aactaatagc tataatcaat atttagttt ggtataggag acaaatcaag	2341
347 tgatctgaac aaattaaatg gttatattt cattgtgaca g ggt tgt gat gga tca	2397
348 Gly Cys Asp Gly Ser 75	
349	
351 gtt ttg ctg aac aac act gat aca ata gaa agc gag caa gat gca ctt	2445
352 Val Leu Leu Asn Asn Thr Asp Thr Ile Glu Ser Glu Gln Asp Ala Leu	
353 80 85 90	
355 cca aat atc aac tca ata aga gga ttg gac gtt gtc aat gac atc aag	2493
356 Pro Asn Ile Asn Ser Ile Arg Gly Leu Asp Val Val Asn Asp Ile Lys	
357 95 100 105 110	

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/08/939,905D

DATE: 11/16/2001  
TIME: 10:20:14

Input Set : A:\Sequence Listing.txt  
Output Set: N:\CRF3\11162001\H939905D.raw

359 aca gcg gtg gaa aat agt tgt cca gac aca gtt tct tgt gct gat att	2541
360 Thr Ala Val Glu Asn Ser Cys Pro Asp Thr Val Ser Cys Ala Asp Ile	
361 115 120 125	
363 ctt gct att gca gct gaa ata gct tct gtt ctg gtaattaata actcctaatt	2594
364 Leu Ala Ile Ala Ala Glu Ile Ala Ser Val Leu	
365 130 135	
367 aattcccaac cattaaaaaq ttgcattgatt ggattcaaaa ttctatggta ttggggttct	2654
369 gatataaatt tgtaattaaa ttgcactaaa aaaaattatc atatactttt aataaaaaaa	2714
371 atttatctaa tttaattttt tattaaaact atttttaaa ttcaatccta actcttttt	2774
373 aatcgagca tgtaagctgg cacccaccgt atatogttgg aagatgctat aaaaccattt	2834
375 aattaatgga tggaatcgt caaaacattt aattcaaaat actcttaattt gtgatttagta	2894
377 atcatgttcg ggcaagttac gttgtgtata attaatttga cttaatcaga taaaaaaaca	2954
379 aatggacgca agccggttgg tatagatatac actggcctgt agaatatgtg gttttcacy	3014
381 tttaaataaa agctagctac tatattat ttagtctttt ttttcttaa acccatttaa	3074
383 cgtgatttt tgactgtgaa acatgtttcc acacacaggc ttagaaactc ctcgcaacta	3134
385 acatctccaa aatttgacta ttattttatg aagataattc atctatgtg ttcaactcta	3194
387 ttatatatat gtatcatcgc agtattaaag attataatag tcaaatatag aagtatatcg	3254
389 ggttaaatgta gttgcattgtg cgacctgtt cgtttaaat gtttattcta tatactttt	3314
391 ttatttgaa aataacgtg aactaaaaac gaaagggtat catatagtt gactttatg	3374
393 ttagagagag acatcttaat ttggcatat gtttataat taattacaat gcatacaca	3434
395 atatttatgc catatctaaa aaatgataaa atatcatagg tataactcaac tatatgat	3494
397 ccccataaca gaaattgtac ttttcttgc gcaatgaact taacatttct gtttgcataa	3554
399 aacaaacatc cacttaaagt gttcaacat atttatgtaa taatttacag gga gga	3610
400 Gly Gly	
403 ggt cca gga tgg cca gtt cca tta gga aga agg gac agc tta aca gca	3658
404 Gly Pro Gly Trp Pro Val Pro Leu Gly Arg Arg Asp Ser Leu Thr Ala	
405 140 145 150 155	
407 aac cga acc ctt gca aat caa aac ctt cca gca cct ttc ttc aac ctc	3706
408 Asn Arg Thr Leu Ala Asn Gln Asn Leu Pro Ala Pro Phe Phe Asn Leu	
409 160 165 170	
411 act caa ctt aaa gct tcc ttt gct gtt caa ggt ctc aac acc ctt gat	3754
412 Thr Gln Leu Lys Ala Ser Phe Ala Val Gln Gly Leu Asn Thr Leu Asp	
413 175 180 185	
415 tta gtt aca ctc tca ggtatacata atcaattttt tatttgcata tagctacaa	3809
416 Leu Val Thr Leu Ser	
417 190	
419 taaaaagtct ctgatacaga catatttgcata taaattaattt tctccataaa cattataat	3869
421 aaaattatca atttatgtac tttaaaaattt tggattgaag ctctttcat ccaacttttta	3929
423 ctaaagttaa ggtgcataataataaactatctct tggattttat taaaagattt	3989
425 aagataagtt aaagtctact tataatcat taatataatgt ata ggt ggt cat acg	4044
426 Gly Gly His Thr	
427 195	
429 ttt gga aga gct cgg tgc agt aca ttc ata aac cga tta tac aac ttc	4092
430 Phe Gly Arg Ala Arg Cys Ser Thr Phe Ile Asn Arg Leu Tyr Asn Phe	
431 200 205 210	
433 agc aac act gga aac cct gat cca act ctg aac aca tac tta gaa	4140
434 Ser Asn Thr Gly Asn Pro Asp Pro Thr Leu Asn Thr Thr Tyr Leu Glu	
435 215 220 225	
437 gta ttg cgt gca aga tgc ccc cag aat gca act ggg gat aac ctc acc	4188

VERIFICATION SUMMARY  
PATENT APPLICATION: US/08/939,905D

DATE: 11/16/2001  
TIME: 10:20:15

Input Set : A:\Sequence Listing.txt  
Output Set: N:\CRF3\11162001\H939905D.raw

L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date